JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# Silent Voice

Preeti Bailke, Yash Munde, Nitin Choudhary, Muaz Sayyed, Ayush Vidhale, Netal Zanwar

Information Technology Department
Vishwakarma Institute of Technology, Pune, 411037, Maharashtra, India

**Abstract** —People who are deaf and mute may have problems communicating with others, and use sign language to express themselves. Communication can be challenging for individuals who do not understand sign language, creating a communication gap. To bridge the communication gap, the person can use a sign language recognizer. The software Silent Voice, a sign language recognizer, converts signs to text, helping in effective communication.

***Keywords* — *Sign Language, Deaf, Mute, Communication***

## I. INTRODUCTION

The primary objective of communication is to transfer information from one individual to another. The act of exchanging, exhibiting, or transmitting information and ideas by writing, speaking, or gesturing is known as communication. In general, people can communicate by talking and conveying their emotions without difficulty. However, deaf and mute people have trouble communicating since they express their ideas through hand gestures, which are unlikely to be understood by a normal person. A need for a sign language interpreter who would convey the thoughts of deaf and mute people to normal people arises. But a sign language interpreter can't always be around, so a replacement is required. A sign language recognizer will convert signs to text. In turn, Silent Voice is a system which uses a webcam to record real-time images of hand gestures with OpenCV. After a picture is captured, image labeling is necessary, and sign recognition is performed using a pre-trained model called SSD Mobile Net v2.

## II. LITERATURE SURVEY

**[1]** The paper introduced a python-based system for sign language recognition. The system classifies 80 words from sign language using two models: YOLOv4 and SVM with MediaPipe. YOLOv4 is used for real-time hand gesture detection, while SVM with linear, polynomial, and RBF kernels is employed for classification. The authors created a dataset of 676 images depicting 80 static signs for training and evaluation. SVM with MediaPipe achieved an accuracy of 98.62%, while YOLOv4 achieved a higher accuracy of 98.8%. **[2]** The research paper described a system for hand gesture recognition using CNN (Convolutional Neural Network) architecture. The system followed a series of preprocessing steps, including grayscale conversion, segmentation using skin masking, and feature extraction using the SURF technique. The preprocessed images were then fed into the CNN layers, which extract features and train the model. The classification phase involved the use of classifiers such as CNN, KNN, Naïve Bayes, Support Vector Machine, and logistic regression. The system aimed to classify hand gestures captured by a camera based on the trained model. **[3]** The research paper focused on the robust modeling of static signs in Indian Sign Language (ISL) recognition using deep learning-based convolutional neural networks (CNN). The study collected 35,000 sign images of 100 static signs from various users. The proposed system achieved high training accuracy of 99.72% on colored images and 99.90% on grayscale images. The system's performance was evaluated using different optimizers and demonstrated effectiveness in terms of precision, recall, and F-score. The research highlighted the superiority of the proposed system compared to earlier works that considered only a few hand signs for recognition. **[4]** An algorithm for real-time recognition of sign language is presented to aid in communication with deaf people. To enhance accuracy, a 3D-CNN combined with optimized dense optical flow is proposed. The collected RGB video stream undergoes optical flow processing and is then fed into the 3D-CNN to extract feature vectors. The system also includes an artificial interaction interface, motion detection module, and hand/head detection module for real-time recognition. Experimental results demonstrate excellent performance and practicality of the proposed system. **[5]** The pipeline combines SSD, 2D CNN, 3DCNN, and LSTM models to capture spatio-temporal dynamics of hands. A new large-scale dataset called RKS-PERSIANSIGN is introduced, containing 10,000 RGB videos of 100 Persian sign words. The proposed model outperforms state-of-the-art models in hand sign language recognition, hand pose estimation, and hand action recognition on multiple datasets. **[6]** To help the deaf and mute communicate, an effective real-time sign language recognition system has been created. Thoughts and ideas are communicated using sign language, a visual language that makes use of body language, hand gestures, and facial expressions. **[7]** It has also been suggested to use thresholding and a skin color model to identify hand gestures, which has applications in robotics and related areas. The YCbCr color space skin color model is utilized to identify and isolate the hand region, while thresholding is applied to differentiate between the foreground and the background. Then, a template-based matching strategy is applied using Principal Component Analysis (PCA). A method for translating sign language into written text has been developed and put into use. **[8]** Hand movements are represented as feature vectors by the Scale-Invariant Feature Transform (SIFT) algorithm. On the edges that are resilient to scaling, rotation, and noise addition, the SIFT features are calculated. **[9]** The paper describes a system for real-time sign language recognition that employs several techniques. The system contains a sign language translation tool, a database-driven method for hand gesture detection, and a CNN-based model for recognition. **[10]** The research also recommends using convolutional neural networks to instantly recognize American Sign Language (ASL). (CNNs). The writers enhanced a pre-trained CNN model using transfer learning to achieve high accuracy in ASL gesture recognition. The system's capacity to handle video input in real-time made it useful for applications like helping the hard-of-hearing and deaf in actual environments.

## III. METHODOLOGY

The technique uses an end-to-end customized object detection model to translate sign language in real-time based on hand gestures. This approach utilizes a mix of Python, Django, HTML, CSS, and JavaScript to implement and execute the strategy.
The Single Shot Detector (SSD), a convolutional neural network architecture renowned for its high speed and accuracy, is used in the object detection process. The two main stages of the SSD are feature map extraction and object detection using convolutional filter applications.

The stages in the suggested methodology are as follows:

1. Collect images for deep learning using the webcam and OpenCV library of Python.
2. Labeling images for sign language detection using the Labeling library in Python.
3. Setting up TensorFlow Object Detection pipeline configuration.
4. Performing transfer learning to train Deep Learning Model.

5. Detecting sign language in real time using OpenCV.

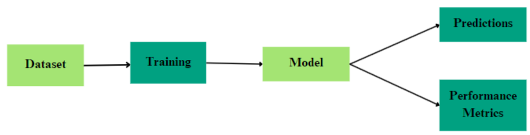


Fig. Predictive Modeling Workflow for Sign Language Recognition

## IV. ALGORITHM USED

- **Convolutional Neural Network:**

A Convolutional Neural Network (CNN) is a specific type of deep learning model that is capable of analyzing input images by assigning learnable weights and biases to different features and objects present in the image. This enables the CNN to accurately distinguish between various elements within the image. Compared to other classification techniques, a CNN needs significantly less preprocessing. While simple methods require filter hand-engineering, CNN can learn these filters/characteristics with the right training. CNNs are artificial multilayer neural networks that can process information from either 2D or 3D data. Each stratum in the network is made up of various planes, some of which can be 2D or 3D. Each plane contains a variety of distinct neuron configurations, where neurons from adjacent layers are connected but those from the same layer are not. By using the right filters, a CNN can accurately record the spatial and temporal components of an image. Furthermore, the architecture did better when fitting to the image collection by lowering the number of parameters and reusing weights. By removing pertinent characteristics from images while keeping crucial data needed for precise prediction, CNN aims to simplify image processing. This is very helpful for developing an architecture that can manage enormous amounts of data in addition to gathering and learning characteristics.

- **Architecture used:**

In order to predict and interpret sign language gestures, a Convolutional Neural Network (CNN) typically consists of three layers: pooling layers, convolutional layers, and fully connected layers. In Fig.1 CNN architecture can be seen. For processing sign language gesture data, these layers are stacked together to create a CNN architecture.

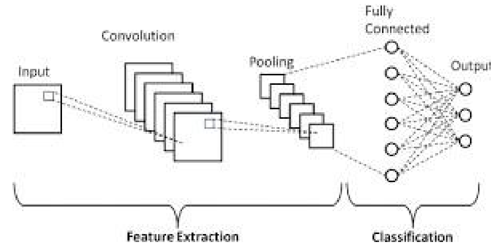


Fig.1 Simple CNN Architecture

By combining convolutional layers for feature extraction, pooling layers for downsampling and spatial invariance, and fully-connected layers for global relationships and prediction, CNN architectures can effectively learn and interpret sign language gestures, enabling the prediction of the spoken words or phrases represented by those gestures.


Fig.2 Sign Symbols ( 1. Thank You, 2. Hello, 3. Yes, No, 5. I Love You)

## V. TOOLS USED

- **TensorFlow:**

TensorFlow is a freely available AI software library that enables developers to construct complex neural networks composed of multiple layers, using data flow diagrams. It was used for classification and prediction.

- **Object Detection API:**

A TensorFlow open-source API for locating and identifying objects in images.

OpenCV:

An open-source, highly optimized Python tool for computer vision issues is called OpenCV. Real-time applications that offer computational efficiency for handling big amounts of data are its main focus. It was used to analyze different real-time images and hand gestures.

- **LabelImg:**

A graphic image annotation tool for labeling object-bounding areas in pictures. The dataset images were labeled with different words such as in Fig 2 respective hand gesture and sign name are shown while in Fig. 3 labeling of image is shown.

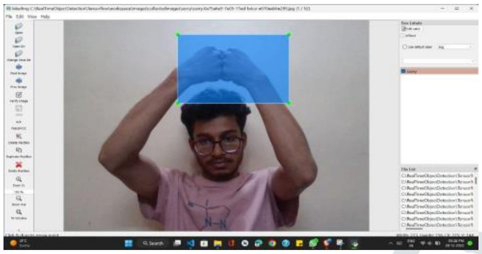

Fig.3 Labeling Image

## VI. RESULTS

A video call system has been made for smooth communication. The output of the Silent Voice system is shown in fig. 4. Sign language has been made by deciding different hand gestures for particular words. For sign language recognition, the dataset was created, and labeled, and some signs were recognized. With the help of the model, the accuracy could be calculated. Accuracy is calculated on the basis of the Confusion Matrix. Confusion Matrix gives the performance of the model.

Confusion Matrix:

Actual values

| | | Positive | Negative |
|---|---|---|---|
| Predicted Values | Positive | TP | FN |
| | Negative | FP | TN |

Where:
TP: True Positive
TN : True Negative
FP: False Positive
FN : False Negative

And with the help of confusion matrix accuracy, prediction, recall , F1 - score is calculated using the below formula:

**Accuracy** = (TP + TN) / (TP + TN + FP + FN )

**Precision** = TP / (TP + FP )

**Recall** = TP / ( TP + FN)

**F1 score** = ( 2* Precision * Recall )/ (Precision + Recall )

Results of the sign recognition model are shown in Table 1 and Table 2.

**Table 1: Model performance analysis**

| No. of Train Images | No. of Test Images | Correct Prediction | Wrong Predictions | Accuracy (%) |
|---|---|---|---|---|
| 120 | 50 | 28 | 22 | 56 |
| 225 | 100 | 66 | 34 | 66 |
| 285 | 150 | 126 | 24 | 84 |
| 350 | 200 | 174 | 26 | 87 |

**Table 2 : Performance Metrics of Model**

| Test Dataset | Accuracy (%) | Precision | Recall | F1 Score |
|---|---|---|---|---|
| Thank You | 87.8 | 0.85 | 0.83 | 0.82 |
| Yes | 89.9 | 0.84 | 0.86 | 0.80 |
| Hello | 93.2 | 0.87 | 0.85 | 0.88 |
| No | 88.6 | 0.87 | 0.82 | 0.85 |

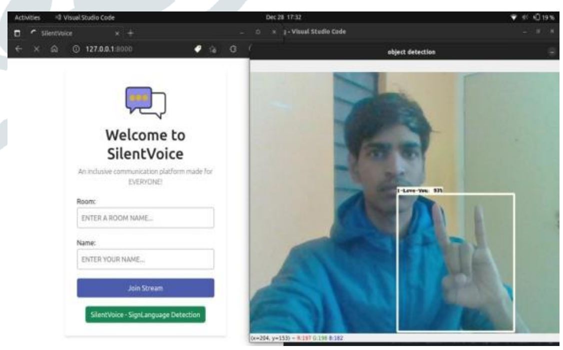


Fig.4 Sign Detection in Real Time

## VII. FUTURE SCOPE

- The dataset could be enhanced by including more images.
- A sign language recognizer could provide a platform for deaf and mute people to express their views in any meeting or online interview. Therefore, it will be

possible in the future to implement this sign language recognition technology in video calls and meetings, which will allow these individuals to excel in their respective industries.

- Accuracy of sign identification must be increased.
- The model can be enhanced by recognizing facial expressions too.

## VIII. CONCLUSION

A sign language detection system's main objective is to offer a workable means of hand gesture-based communication between normal and deaf individuals. The suggested system is accessible through a webcam or any built-in camera that recognizes and processes signs. The model's findings lead to the conclusion that, in the presence of controlled light and intensity, the suggested system can identify gestures correctly. Custom gestures can also be readily added, and the model will be more accurate if there are more photos taken at various angles and frames. As a consequence, the model can be quickly expanded in scope by growing the dataset. The model has some drawbacks, such as environmental factors that decrease detection, like low light intensity and uncontrolled background.

## REFERENCES


[1] R Sreemathy, MP Turuk, S chaudhary, K Lavate, A Ushire, S Khurana , “ Continuous word level sign language recognition using an expert system based on machine learning”, International Journal of Cognitive Computing in Engineering , 2023, 170 - 178.

[2] Akriti Goyal, Deepanshu Dhar, Paras A Nair, Chirag Saini, Supreetha S M, Chetana Prakash, Sign Language Recognition System, International Journal of Engineering Research & Technology (IJERT) ,ISSN: 2278-0181, ICEI - 2022 Conference Proceedings

[3] Ankita Wadhwan, Prateek Kumar, “ Deep learning-based sign language recognition system for static signs”, in Neural Computing and Applications, 2020

[4] Kai Zha1, Kejun Zhang, Yu Zhai, Daotong Wang, Jianbo Su, “Real-Time Sign Language Recognition Based on Video Stream” - 39th Chinese Control Conference
July 27-29, 2020, Shenyang, China

[5] Razieh Rastgoo a, Kourosh Kiani a, Sergio Escalera, “Hand sign language recognition using multi-view hand skeleton” - 15 July 2020

[6] R. Rumana, R. Sandhya Rani, and R. Prema, "A Review Paper on Sign Language Recognition for The Deaf and Dumb," in Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2017, pp. 400-403.

[7] M. K. Ahuja and A. Singh, "Hand Gesture Recognition Using PCA," in Proceedings of the 2015 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), Kozhikode, India, 2015, pp. 1-5.

[8] A. Kumar, A. Ahmad, A. Jaiswal, and A. Kumar, "Hand gestures recognition conversion to speech," in Proceedings of the 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, India, 2018, pp. 751-756.

[9] A. Pathak, A. Kumar, P. Priyam, P. Priyanshu, and G. Chung, "Real-Time Sign Language Detection," in Proceedings of the 2017 3rd International Conference on Advances in Computing, Communication & Automation (ICACCA), Dehradun, India, 2017, pp. 1-5.

[10] B. Garcia and S. Alarcon Viesca, "Real-time American Sign Language Recognition with Convolutional Neural Networks," in Proceedings of the 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, 2018, pp. 1487-1492.